DOCUMENT RESUME

ED 191 865 TM 800 477

AUTHOR Davis, John N.; Moriel, James, Jr.
TITLE Using the Rasch Model to Establish Equivalent Certification Tests.
PUB DATE Apr 80
NOTE 49p.; Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980). Best copy available.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Basic Skills; Criterion Referenced Tests; Cutting Scores; *Difficulty Level; *Graduation Requirements; High Schools; Item Analysis; Item Banks; Mathematical Models; *Minimum Competency Testing; Reading Achievement; *Scaling; Secondary School Mathematics; *Test Construction; *Test Items; Writing Skills
IDENTIFIERS *Rasch Model; Rasch Scaled Scores; San Jose Unified School District CA

ABSTRACT

In order to compose equivalent certification tests in the basic skills in reading, mathematics calculations and applications, and writing for high school graduation requirements, the San Jose Unified School District, California, used the Rasch computer program to arrive at the scaled difficulty level of test items. The procedure for achieving a Rasch scaled item bank included: (1) production of Rasch unit scale values (RIT); (2) determination of the difficulty index of each competency; (3) production of the average difficulty of each performance indicator (PI) in RITs. Equivalent tests can be constructed from the scaled item bank using RIT values or difficulty levels for each PI and/or competency which average the same as the first tests. A degree of flexibility can be built in by using scaled items of RIT values such that their average is equal to the desired level. In addition, equal cutting scores on tests are maintained. Whatever the quality or criterion of the first test, the second will be nearly an equivalent version of it. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED191865

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Using The Rasch Model to Establish Equivalent Certification Tests

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. Davis
J. Moriel
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

John N. Davis
James Moriel, Jr.

San Jose Unified School District
San Jose, California

Paper presented at Joint Session 15.05: Advances in The Application of the Rasch Model to Practical School Problems; national conference of the American Educational Research Association and the National Council on Measurement in Education, Boston, April 1980

TM 800477

The use of the analytical processes for item characteristic curves in applications to criterion referenced test items used to compose Certification Tests in Basic Skills for high school graduation requirements is described. The use of the model is part of the item analysis process developed by the Portland Oregon, Public Schools. Its applications feature reasonable restrictions for item selection so that equivalent test forms and equal cutting scores for graduation certification in Reading and Mathematics Calculations and Applications are maintained.

Using the Rasch Model to Establish Equivalent Certification Tests

John N. Davis
James Moriel, Jr.
San Jose Unified School District

1.0 Introduction

The concerns felt in the various communities served by public schools for a guaranteed level of performance of students in the use of the basic skills of reading and writing in English and computing and applying the mathematics knowledge to life situation are clearly evident in this country. One obvious result is the recent sets of legal or administrative requirements in the several states to monitor student achievement in these learning areas and to report the results individually to the parents and students. Two states, Florida and California, have state laws requiring a student to pass certification requirements of proficiency in these learning areas before a high school diploma can be granted.

Even though the implementation of the Florida requirement has been delayed until 1982 due to court action (case: Debra P. vs Turlington 1979-80) the challenge to the state developed testing process was not upheld. The same test was used across the state, provided by the Florida State Department of Education. However, in California's situation, the legislation (Hart: AB3408: 1976 and AB65, 1978) requires that this process of certification and assessment be developed by local school district boards of education. Laws require that these locally adopted standards of proficiency be defined for these competencies in these basic skills.

This legislation specifically disallows the use of state-wide testing or the use of grade equivalent scores or other comparisons between stduents on this achievement in these areas as a means of establishing these standards. Thus, the need exists to establish local standards of success.

Since tests are the most practical method to assess achievement for the large majority of the students, the standards translated to cutting scores on tests are the means used to identify those students who have met the required proficiency levels. Those who pass can then be certified for graduation after June 30, 1980, provided they have also met specific course and unit requirements established by the state and the local school boards. The California laws also require a monitoring

of student progress in these learning areas with assessments given at least once in grades 4 and 6, once in grades 7 and 8 or 9, and twice in grades 9 through 12.

Finally, there must be levels of performances (cutting scores) established as expected proficiency levels for the grades below 10 so an individual student's progress can be monitored. Then target teaching can be employed to increase the students' operational knowledge and achievement in these areas before the graduation certification requirements can be met.

This process has resulted in the adoption of standards which define the specific competencies clearly in these learning areas. Competencies are usually stated as broad curricular goals with a subset of teaching-learning objectives identified with each curricular area. These subsets are frequently labeled as individual student objectives and called performance indicators. These performance indicators refer to specific content knowledge or its application which the student is expected to achieve. They are stated in terms of what each student "can do" to meet the specific performance requirements. The aggregation of the student's success on passing items directly mapped to these curricular blue prints then becomes the means of determining acceptable student performance on the competency defined in a curricular sub area. Summing this success over the items and across the competencies and their performance indicators provides the means of determining acceptable performance of a student in each of these curricular sub-areas. The cutting scores then have a major role to play in the process.

2.0 Local District Action

One example of a set of such statements of Basic Skill Competencies are those adopted for graduation in the San Jose Unified School District, San Jose, California (SJUSD, 1978; refer to appendix A for statements.) The competencies, their performance indicators, and the corresponding of sets of the appropriate test items for the performance indicators are definitions of curricular domain segments in these basic skills. When the domain segments are defined within adopted statements of curricular courses and sequences so they become parts of the official district, county, or state curricular specifications, they certainly have the administrative validity required in the legal intent and local expectation. Further, these blueprints or frameworks with the mapped test items provide a logical, practical structure upon which valid, reliable tests can be developed and instructional processes targeted. Concerns raised by McClung and others (1977) about curricular validity of

the certification processes and the necessary instructional validity of the teaching acts related to helping students meet such standards are more readily addressed.

3.0 Specification of Test Content

The processes followed in the San Jose Unified School District have targeted on developing such domains. The statements of those competencies and their performance indicators were jointly developed by a special task force of parents, teachers, students and administrative personnel in 1977 and 1978. Attention was also paid to the official district curriculum, to those of other Santa Clara County school districts, and to the state curricular frameworks. When the local board of education adopted the statements as those competencies required for graduation certification and the performance indicators as the set of standards against which a student's proficiency would be assessed, the basis for determining content validity of tests and other assessment processes was established.

Content and forms of the statements developed thus have "local ownership" and wide common interpretation and understanding across the necessary representative members of the district. They are not as concise as those which can be stated in parallel fashion in the subject fields. For example, all of the competencies in mathematics could be replaced by requiring a student to be proficient in the knowledge of and the use of rational numbers in their operations and applications in school settings and functional transfer or life applications. However, the use of operational definitions of competencies and performance indicators, with understanding and acceptance by students, parents and the general public act to provide a more reasonable basis for determining processes, including tests and the curriculum taught.

During the period this competency framework was being developed, the district cooperated with ten other districts in Santa Clara County to develop a pool of items that were mapped to nearly equivalent performance indicators. Large field trials of these items were conducted out across samples of 19,000 students in Grades 9 and 10 using 44 high schools in this suburban-urban area. Analysis of the items in terms of their response patterns, success ratios, and these results for different achievement groups were conducted. The items were screened by professional groups under a contract of services with Dr. Carmen Findley, American Institute for Research, and with the involvement of teaching and research personnel from the participating districts. Interim results of the field trial tests were given to the district task force establishing the competency framework so that both content and item types were known, analyzed, and screened for their appropriateness to the tasks.

Some writers as Zieby and Fremer (1979) have begun to describe the necessity for such careful processes in defining content of the tests and these actions.

After re-writing some items, excluding others, and verifying the content validity of all of those acceptable as meeting the standards of good test requirements, a common county item pool was developed. These items were now mapped to the local district competency framework (see appendix A) and then were formatted into a second set of field trial tests given in the fall of 1978. Students in grades 7 through 12 of the district were tested. Representative samples, of 350 students were tested in grades 7, 11 and 12; over 2500 students in each of grades 8, 9 and 10 were tested in each curricular area of writing, reading, and mathematics calculations and applications. These tests involved more than 90% of the faculty members in schools enrolling the grades 7 through 12 so information as to test appropriateness to the competency framework and to the curriculum of the district was developed and analyzed. The specific content of the tests was thus widely disseminated in each school through the faculties and their student groups who were tested, particularly those who were to be affected by the graduation certification requirements.

4.0 Establishing the cutting scores

This knowledge of the test content was developed with the task force so judgements as to the minimum acceptable levels of performance could be evolved. After discussion of the first set of field trial data, the districts history of student performance on sub-scales of group tests, and comparisons with student results as reported for groups by the National Assessment of Educational Progress (Mullis, et al, 1977), the second set of tests were evaluated by the task force members. Since tests in reading and mathematics were clearly targeted to the competency framework and were composed of items mapped to these criteria, Nedelsky's method of judgement was applied to the mathematics test forms and Angoff's method was applied to the reading test forms. These processes as summarized in the Manual for Setting Standards on the Basic Skills Assessment Tests, ETS (Zieby and Livingston, 1977) were followed.

These probablistic judgements of success of those students judged to be minimally qualified in satisfying the performance indicators, were used with these 38 representatives on the task force acting as judges. Additional meetings of parent and student groups were held in three separate representative high school attendance areas to

provide additional judgement data in the test forms for the task force to consider with the Angoff method of assigning probablistic values.

The results on the second field trials, using the same test forms judged by these groups, were then given to the groups for consideration.

The task force also examined the holistic scoring methods used on the four tasks required in the writing sample and considered the student results on that set of field trial tests. Since the pattern of judging tests had been established by the work with the Nedelsky and the Angoff methods on the criterion referenced mathematics and reading tests, the task force followed the logic, but not this complete analysis of the several sub-sets of check-lists used in the historic process and related analyses for papers which were judged as failures.

Their recommendations of cutting scores of 65% in mathematics calculations and applications (of 88 items), 70% in reading (of 50 items) and 75% in writing (3 of 4 tasks) were given to the local board of education and adopted. A further requirement of success on each competency was also added in the process to ensure that a student demonstrates a breath of achievement across the competency framework; this requirement specifies that he pass a majority of each of the items mapped to each competency.

Thus both the definition of the Basic Skill content, the level of proficiency required and the extent of this achievement were established. Further these field trial tests then became the models and in effect, the specifications against which the formal required testing had to be developed to maintain these same cutting scores since that date. There have been three series of testing carried out since 1978; each test in the competency area has been mapped to the same competency framework with the requirement of maintinaing the cutting scores.

5.0 Use of the Rasch Model

This need to establish equivalent tests to correspond to these model specifications led to investigations of the appropriateness of the Rasch model to provide items which had the same characteristics of those used in setting the standards and tested in the district student populations. The writings of Wright (1977), Ryan and Hamm (1978) and discussions of the developments in Latent Trait Theory (Hambleton, et al 1978) were examined as to the rationale for the Rasch model and its use in the local process. Particular focus following the linking networks developed in Portland, Oregon by Ingebo and Forster (1977, 1979) was given to the relationship of items shared on the various forms used in the field trials so the item pool which was mapped to the competency framework could be characterized as

an item bank.

The experiences described by Choppin (1977) in developing item banks used in England and Wales were examined in terms of the application of the Rasch concepts. The need for the best fit of items to criterion specifications as defined by the performance indicators and to the success criteria as specified in the cutting scores led to the acceptance of the Rasch model as the best method of item analyses and determination of the item difficulty when coupled with the traditional approaches of item analyses.

The use of the item characteristic curve analysis as developed by Forster in Portland, Oregon to plot actual student success adds a necessary component to considerations of point biserials, success ratios of "p" values, and item analyses by distractors for high and low scoring students needed to "engineer" equivalent tests with these criteria.

Two observations appear to be appropriate at this point. First, the ability to arrange a set of items specified with a relative difficulty to a performance indicator or student learning objective provides a clearer, more precise definition of a subset of a curricular domain. The field of mathematics has established careful definitions of the concepts of domain, range, function and relation. The position can now be stated that there are relations now specifiable in these basic skill curricular areas which have a collection of numbers assigned to test items (RITS or Rasch values), that these numbers and their corresponding test items are the domain of definition of the relation; that the relation corresponds to the performance indicator and thus is rule which assigns the item and its difficulty value to the curricular map or framework. The range of the rule is the set of items and their corresponding numeric difficulty values; the domain is defined in the rule.

It follows that the rule (relation) is not a function as defined in mathematics (Begle, 1954) since there are demonstrateably many items with different associated values which correspond to each performance indicator and no one-to-one correspondence exists between each performance indicator and a single item. The point remains, however, that it is now possible to be more precise about definitions of curricular domains and the relations or rules which can be defined within them so examples of required student tasks, the items, have corresponding mathematical values.

6.0 Engineering equivalent tests

Next, the process of "engineering" a test to a criterion of pass or no pass has become more precise. The use of the Rasch calibrations of test items and the accompanying analyses of (1) item curve plots of groups of student behavior and (2) the student results on the entire test provide a set of procedures which result in equivalent tests as to content and as to reliability. The item fits using the Rasch model also provide means for diagnosis of progress of groups of students on these items in the item bank.

Restricting the use of items on these certification tests to those which have a definable effective range, on the horizontal difficulty scale, is another means of improving the accuracy of an item, and in sum, that of the accuracy of a test. Since there is no present curricular definition of what is termed the item information function (Birnbaum, 1968; Hambelton and Cook, 1977) it becomes more practical to use this approach of restricting item composition of a test to solve the problems of using items for their discrimination ability. The use of the Rasch analyses separately establishes difficulty values determined by those students who have a 50% chance of passing the item, e.g. where ability of the student equals the item difficulty. This concept is quite tenable to persons involved in judging tests for minimum competency levels needed for certification.

Further, the processes used in establishing the cutting scores, the criteria for passing on total tests, did not consider the concept of weighting some items as more valuable than others, although it is theoretically feasible to consider such an approcach. The assignment of equal weights to items which have been judged to be curricularly valuable and mapped under the relations between the performance indicators and their sets of corresponding items has more relevance to the problem of maintaining equivalent tests and their matching equivalent cutting socres.

Another concern frequently expressed on a theoretical base is that of adjusting the Rasch or other mathematical models used in item response analysis for supposed guessing on the part of students. The problem may deservedly require theoretical treatment, if in fact, there is research evidence that certain patterns do exist in student data so that the randomness of item response choice can be isolated and analyzed. Another approach which can reasonably be used with the one parameter Rasch model is to examine the plotted performance curve of student responses on

an item, and in particular for those students who have less than 25% success on the item. Evidence already reported by studies in the Northwest Evaluation Association (NWEA, Forster and Ingebo 1977, 78, 79) indicate that the best fit of an item for effective measurement occurs where the shape of the curve approximates that of the secant at its point of inflection. Eliminating the use of items which have curves that do not approximate this ideal is in order.

"Engineering" a test by excluding such items which have performance plots which display large dispersions of student results at the lower left segment of the logistic curve follows analogous practice in industrial engineering where the performance data used to develop theoretical performance curves are again used to restrict the use of the curves to practical applications. Dangers of extrapolation are well known.

Considerations of item response patterns of students on these tests reinforce the working hypothesis of these writers that students select from among two or three possible distractors when they are not certain of the correct response, rather than randomly selecting an answer. This type of data can then be used to rewrite the question so the patterns fall into those which can reasonably be used to diagnose the learning problems, then the effort of corrective teaching can be addressed to help the student in meeting the criterion of success on the item and thus the general criterion of success on the test, that of the cutting scores.

7.0 Maintenance of cutting scores using the Rasch model

Since the Rasch difficulty values of the items do not change when students are tested in the same Basic Skills in these grades 9 through 11, it is possible to maintain the same difficulty values for each domain subset reflected on the tests which existed in the prototype used to establish the cutting scores. The problem now has at least one solution in that when new items are selected to match the performance indicators, these replacement items can be matched with the same RIT value. Alternately it is possible to maintain the same average Rasch difficulty value over a performance indicator (the relation rule which maps the items in its domain). Since the cutting scores were established as a proportion (percentage) of the original sets of domains to be mastered, the maintenance of comparable content of equal Rasch difficulty on each of the succeeding tests is now assured. In this way the relative effect and purposes of the cutting scores are continued.

8.0 Statement of procedures utilized with cooperating California districts

In order to simplify this rationale and these procedures of explaining this process of developing equivalent tests to colleagues working in other California school districts, a brief statement of procedures has been developed for such discussions with them. It is included here as an example of one means of approaching the problem of developing equivalent competency tests using the Rasch model. The same items can then be utilized by using cooperating districts, each developing their sets of equivalent tests following their own specifications as defined by their tests which they used to establish their cutting scores, the criteria of passing the certification competencies.

How to Develop Equivalent Competency Tests
- Using the Rasch Measurement Model -

When you administered your first "official" battery of competency tests in your district, you carefully followed a process in test item selection and test construction. Your tests should have been instructionally and curricularly valid. The level of difficulty was based upon your field testing and criterion for passing.

We would like to share our own experiences with you so we can use them as common grounds for discussion. There may be small differences in our approach or in emphasis from yours, but let's agree to start discussion from our past work. As the result of field tests, we anticipated negatively skewed distributions for real tests. We know that normal distribution curves would have failed fifty percent of our students. The "p" values selected were therefore largely in the 0.65 - 0.85 range. This "p" value range would pass about 75 percent of our population of 9th and 10th grade students. This "p" value range seemed to be reasonable against our field test experience and reinforced the expectations of our citizens and professional groups in light of the student's first attempt at passing their competency requirements. Various discrimination indices were used to check the test items, such as point biserials of 0.3 or better, and proportions of students passing items categorized against our curricular maps of competencies.

In curricular areas, we carefully reviewed the item analyses and item characteristic curves to see where our students were weak; that is, to highlight the skill areas which needed an additional focus of instruction. Ostensibly, we were as conscientious as possible in designing and formating our tests. In the technical areas of test production, we wrestled with the problem of clear graphics for those test items that needed them; we were also careful about exactness of typing and reproduction quality. We were careful in this endeavor since we did not want to bias our field trial data or to make changes to test items by careless reproduction.

We believe our tests were not too long, were curricularly fair, and were reasonable in terms of instructional validity. For instance, the question as to how much time should be allowed for each test was addressed and implemented. As references we examined situations with strict time frames; but we felt that others without these were more accommodating to our needs. In particular we

think the best analogy to use is that followed in the State Driver's Licensing process. When the Department of Motor Vehicles begins to time us all on their written Drivers' Tests to determine minimally qualified drivers, then we believe that is when we should hold fast to a limited time frame for student tests of minimum competency. In essence, we maintain that tests of minimum competency should not be strictly timed, but held within certain practical parameters. When the student hasn't finished and the allotted time has run out, we allow more time at an arranged hour and place and so the student can say "I really can't do anymore." Then we should document the steps which have been taken to that effect. It's unfortunate but necessary to record these efforts taken to protect the students as well as the school district under requirements of law in California.

The initial competency testing phase was over with the school year 1978-79 and now more decisions have arrived with the next phases. They involve several choices: Do we re-test students using entire test batteries for each learning area, such as Reading and/or do we test only for those competencies not achieved? Answering these questions positively could develop a two phased testing program. The first phase requires full batteries of testing, and the other phase for individual competencies with tests targeted to each of them. In either case, developing equivalent measures is needed for each phase.

What do we mean by an equivalent measure? We define an equivalent test as one that measures the same skills and at the same difficulty level as the first test that was used. To test otherwise is to raise a challenge of the concept of fairness. Moreover, if we don't use different items measuring the same intent, our tests will also be unfair. Can we ensure that the "p" values for items will hold from one population of students with one time period to another set of students at a later date? What can we do to avoid as many problems as possible?

We will delegate the tasks of comparability of content of alternate items to our curriculum persons, but offer our help in this process as well. We will now turn to the question of relative difficulty of items within these curriculum maps.

We have a recently added tool in test measurement to help us; this is the Rasch Measurement Model. This mathematical model is population and test free. It produces an equal interval scale value that is independent of the population tested or the arrangement of the items within a test. It is based upon a logical concept embodied in the Latent Trait Learning Theory. Essentially,

it rests on the principle that more able students will get more difficult items right than less able students do. The Rasch formula expresses this function - ability of the student versus difficulty of the item - using the natural log e. The probability that a given student will get an item correct is 50% if the item difficulty is at the student's ability level, with regard to this learning trait.

The formula is:

$$p = \frac{e^{(a-d)}}{1 + e^{(a-d)}} \qquad \text{and } e^{(o)} = 1$$

If ability (a) equals difficulty (d) then we can see that it will produce $p = \frac{1}{2}$ or 50%.

We have learned that adequately large populations (at least 250 students per item) should be used to field test items using the Rasch computer based system to arrive at the scaled difficulty level of the items. These difficulty indices are needed to develop the equivalent tests. A school district with a Rasch scaled item bank can easily design, format and produce equivalent test instruments. The procedure we use for achieving this goal is relatively simple:

1. Our first "official" test was developed from a Rasch scaled item bank which had been processed through our Rasch computer program to produce RIT (Rasch Unit) scale values. (This program also generates traditional data for us to use: biserial correlations, proportions of success, and lists of student choice by item discriminators).
2. Now we can determine the difficulty index (RIT value) of each competency. The procedure we use is: take the scaled RIT value for each item measuring a given competency, add the RIT values, divide by the number of items used, and we have completed the average (RIT) difficulty for the competency. This gives us our experimental base to work with curricular personnel for their actions in selecting items for their next test subscale.
3. If you also used all or most of your Performance Indicators (P.I.) with a certain number of items for each P.I., as we did, then we would then follow the same procedure as above. This process produces the average difficulty of each P.I. in RITs. We think that working with defineable sub sets of curriculum by P.I. gives us a more valid claim to our having equivalent tests. It also provides separate samples of each skill for comparison of students, programs and schools.

If you agree with us, then your equivalent tests can be constructed from your scaled item bank using RIT values or difficulty levels for each P.I. and/or

competency which average the same as your first tests. A degree of flexibility can be built - in by using scaled items of RIT values such that their average is equal to the desired level.

Example: Competency (N) Performance Indicator (n)

C_N P_n

C_N = RIT 195

P_1 = 190 3 items: a, b, c
(Item values)
a) 185
b) 190
c) 195

P_2 = 195 3 items: a, b, c
a) 189
b) 195
c) 201

P_3 = 200 3 items: a, b, c
a) 196
b) 200
c) 204

The referenced N = RIT 195 means that the average RIT value or scale score was made up of 3 P.I.'s each having 3 items or a toal of 9 items measuring the one competency.

A. Measurement by Competency ONLY
Here you can use 9 items of reasonable RIT values such that the average equals 195 plus or minus a standard error of measure.

B. Measurement by Competency and Performance Indicator
In this case you look at each P.I. and select items for it from the item bank that will average the RIT value of the first test. By doing this carefully, the average RIT for the Competency will also be achieved.

C. An iterative process may be necessary for A or B above but it is easily achieved.

Your end product will only be as good as your original test. At least, whatever the quality or criterion of the first test, the second will be nearly an equivalent version of it.

REPORT AM6R40
RASCH

SAN JOSE UNIFIED SCHOOL DISTRICT
RASCH COMPUTER SYSTEM

PAGE 28
3/17/79

05010CSETS COMPETENCY TEST
00018 000018 RIT= 223.4 PCT= 34.0 PT.BIS=0.28 MN .SQ.FIT= 1.53

2,350 Students tested (Grades 7-12)

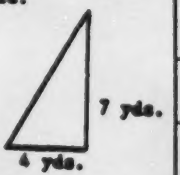
PERCENT GETTING THIS ONE CORRECT

100
90
80
70
60
50
40
30
20
10
0

150. 160. 170. 180. 190. 200. 210. 220. 230. 240. 250.

SCALED ACHIEVEMENT

18. Find the area of this triangle.



1. 11 square yards
2. 14 square yards
3. 28 square yards
4. 56 square yards

Figure 1A

05010CBETS COMPETENCY TEST GRADE = 08
00010 00008 RIT= 223.3 PCT= 29.2 PT.BIS=0.32 HN .50.FIT= 0.33

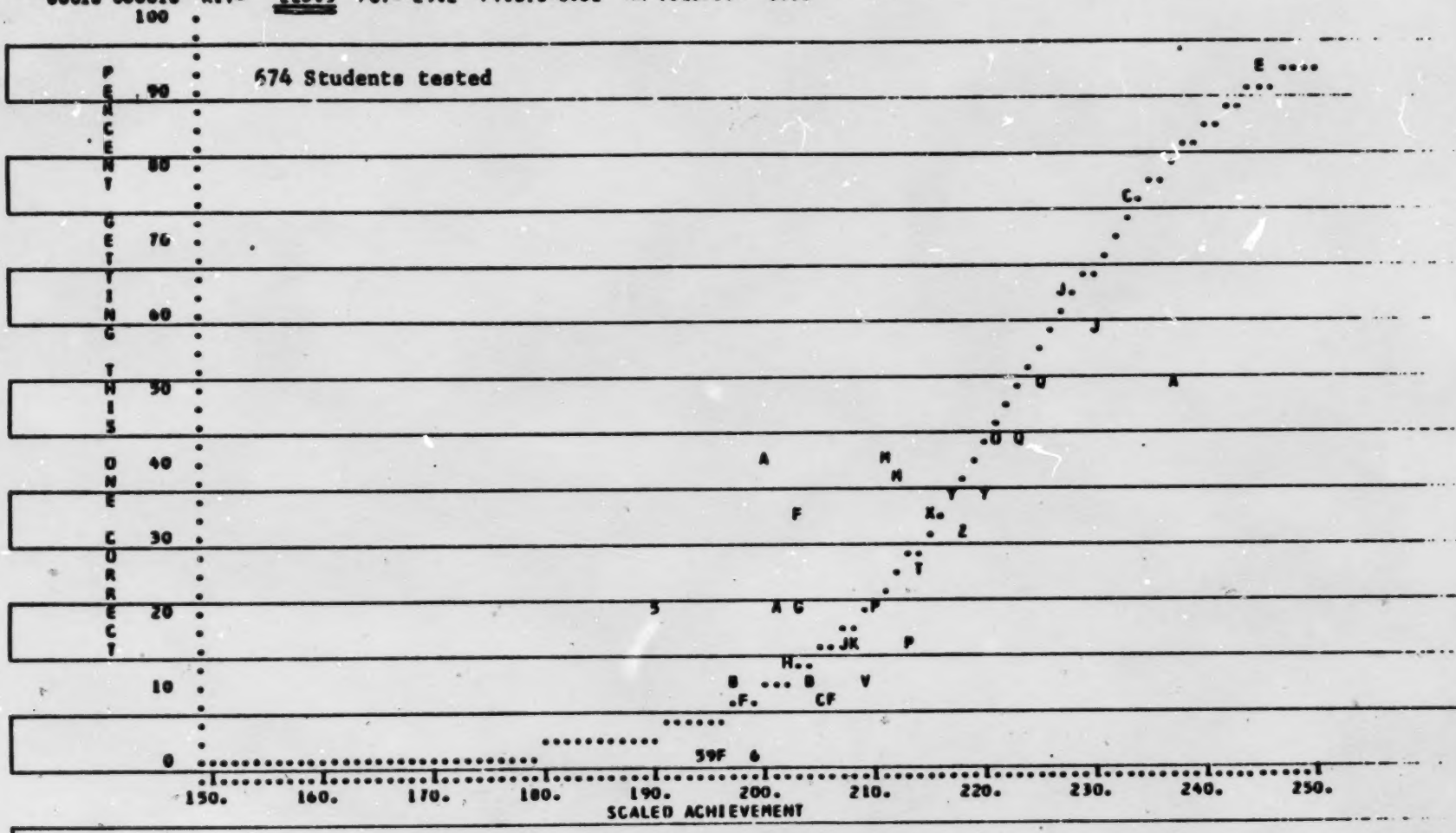


Figure 1B

0901OCBETS COMPETENCY TEST GRADE = 09
00018 000018 RIT= 223.1 PCT= 34.8 PT.BIS=0.40 MN .SD.FIT= 0.29

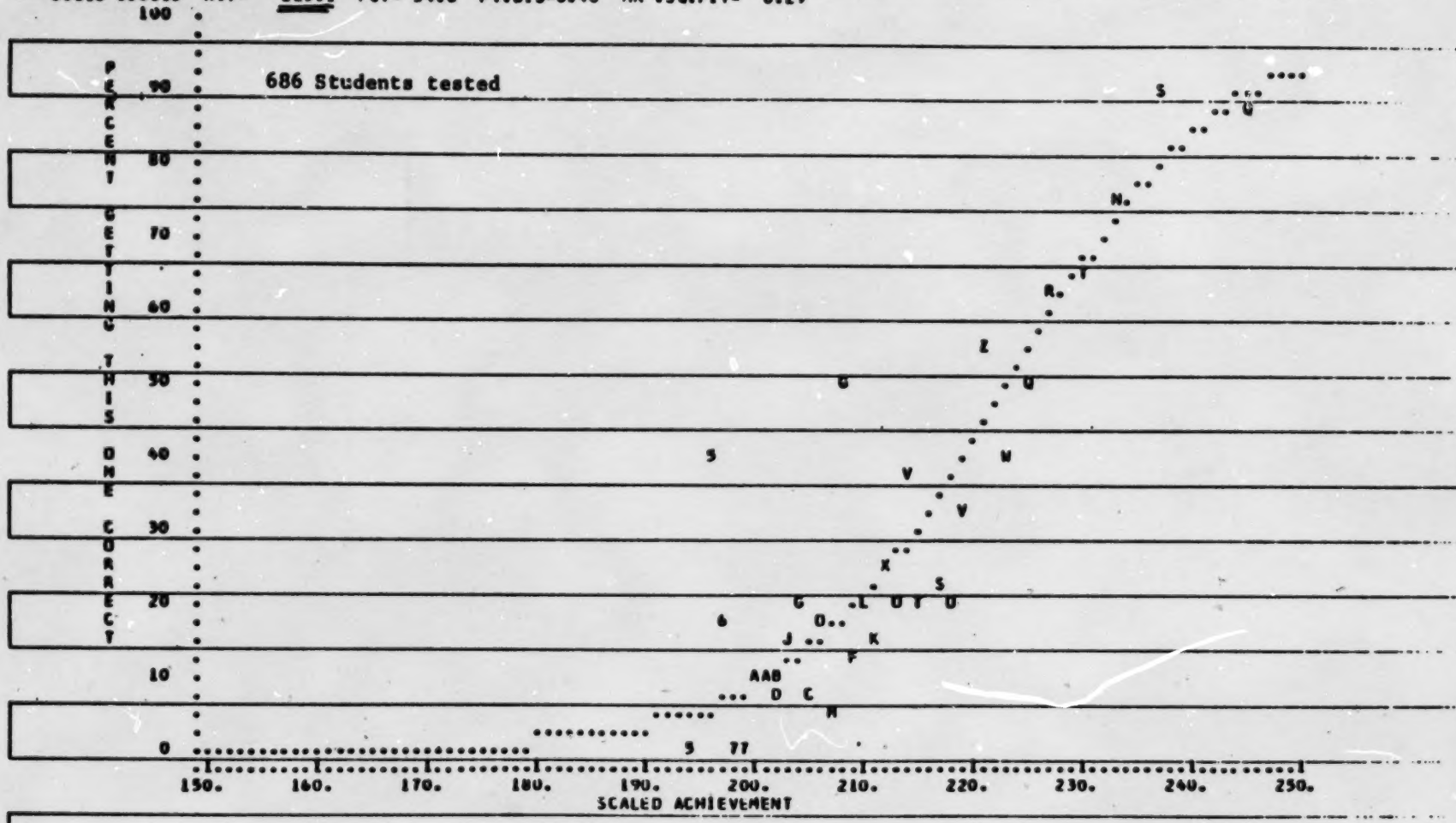


Figure 1C

05010CDETS COMPETENCY TEST GRADE = 10
00018 0J0018 MIT= 221.9 PCT= 38.2 PT.BIS=0.12 MN .SQ.FIT= 5.48

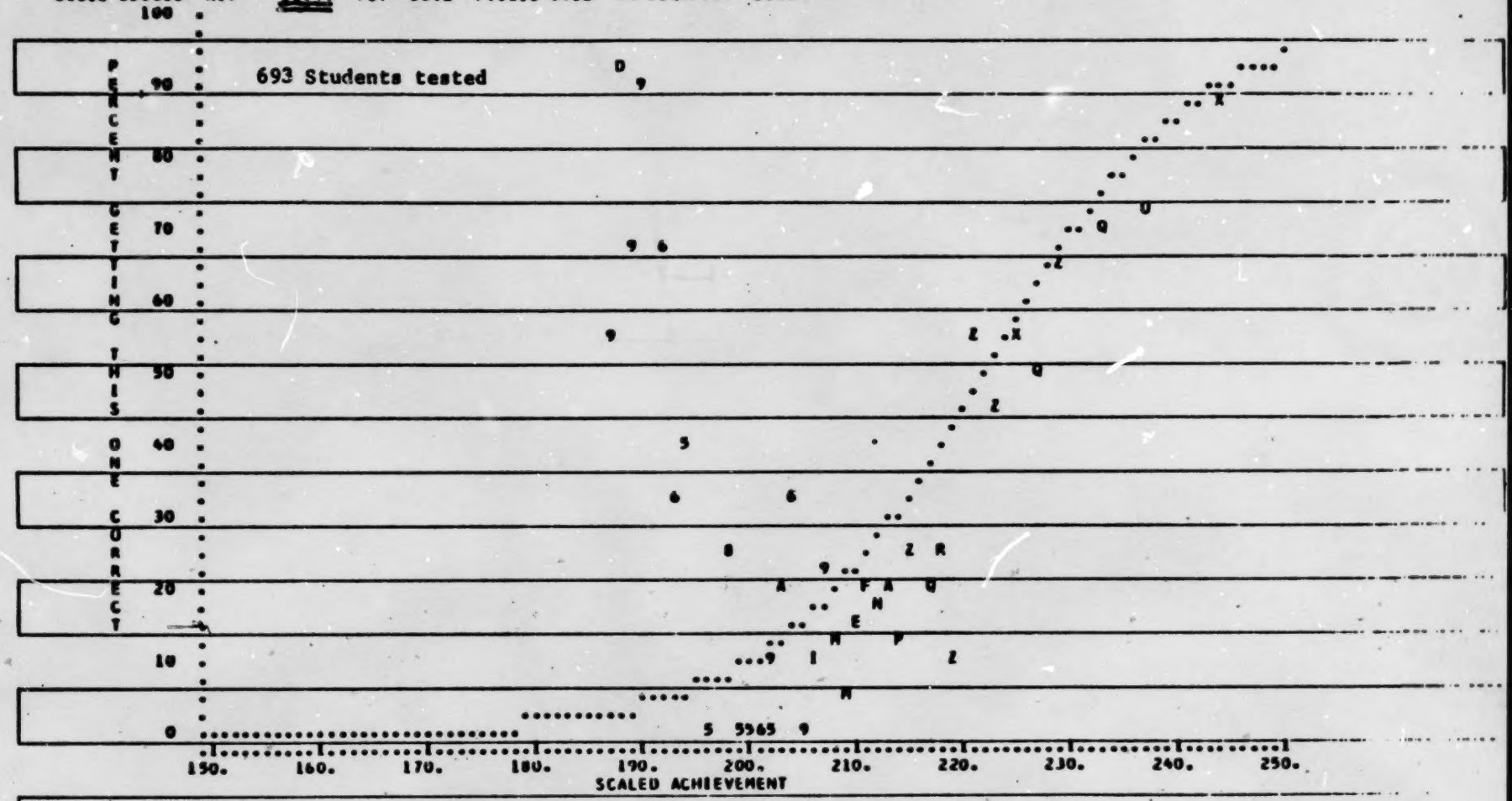


Figure 1D

OSOIOCBETS COMPETENCY TEST
00022 000022 RIT= 184.4 PCT= 92.4 PT.BIS=0.44 MN .SQ.FIT= 0.13

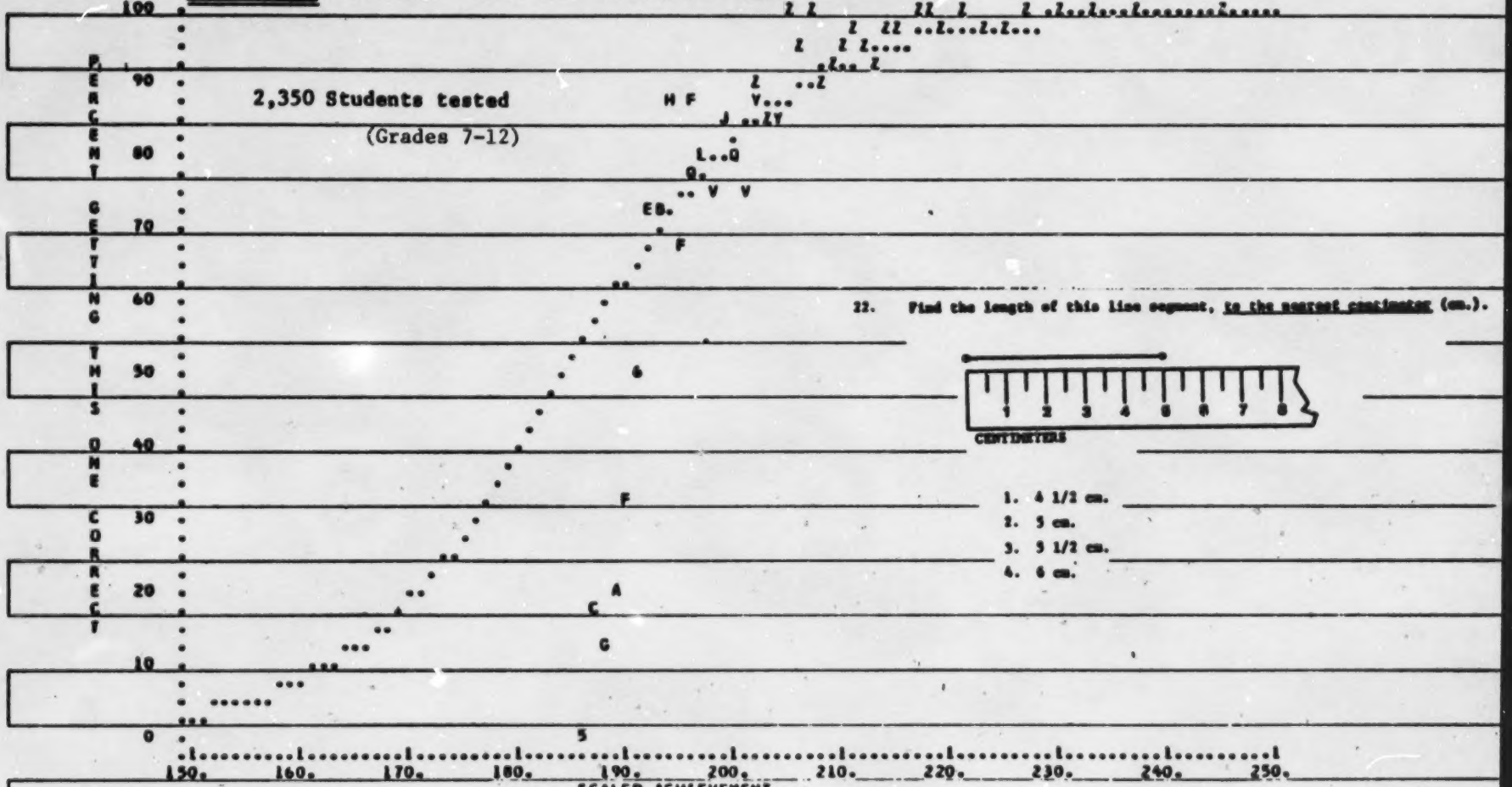


Figure 2A

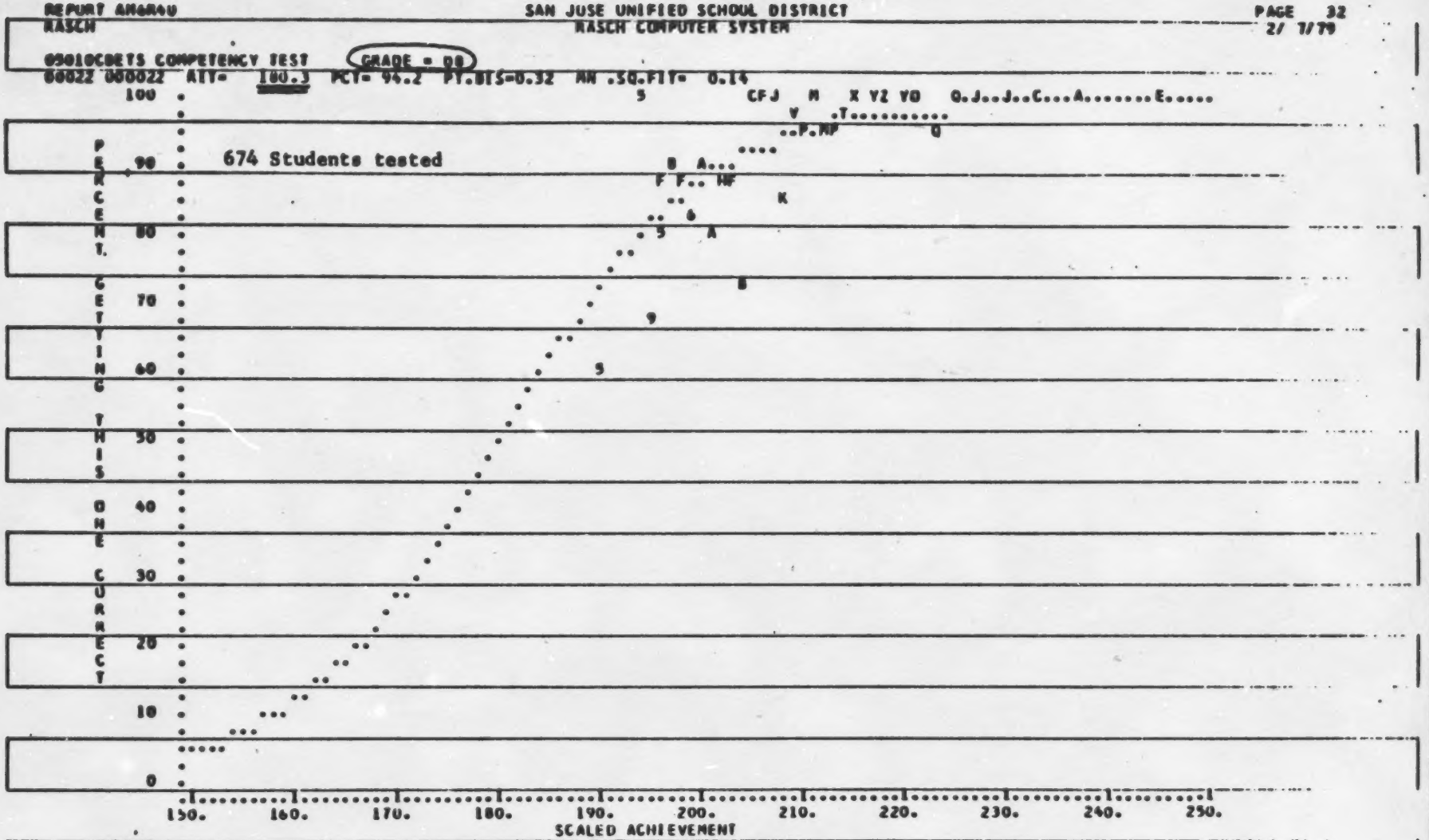

05010CBETS COMPETENCY TEST GRADE = 08
00022 000022 ATT= 180.3 PCT= 94.2 PT.BIS=0.32 AN .SQ.FIT= 0.14
674 Students tested
PERCENT GETTING THIS ONE CORRECT
100
90
80
70
60
50
40
30
20
10
0
150. 160. 170. 180. 190. 200. 210. 220. 230. 240. 250.
SCALED ACHIEVEMENT

Figure 2B

0501OCBETS COMPETENCY TEST (GRADE = 09)
00022 000022 AIT= 184.1 PCT= 93.3 PT.BIS=0.35 AN .SD.FIT= 0.14

686 Students tested

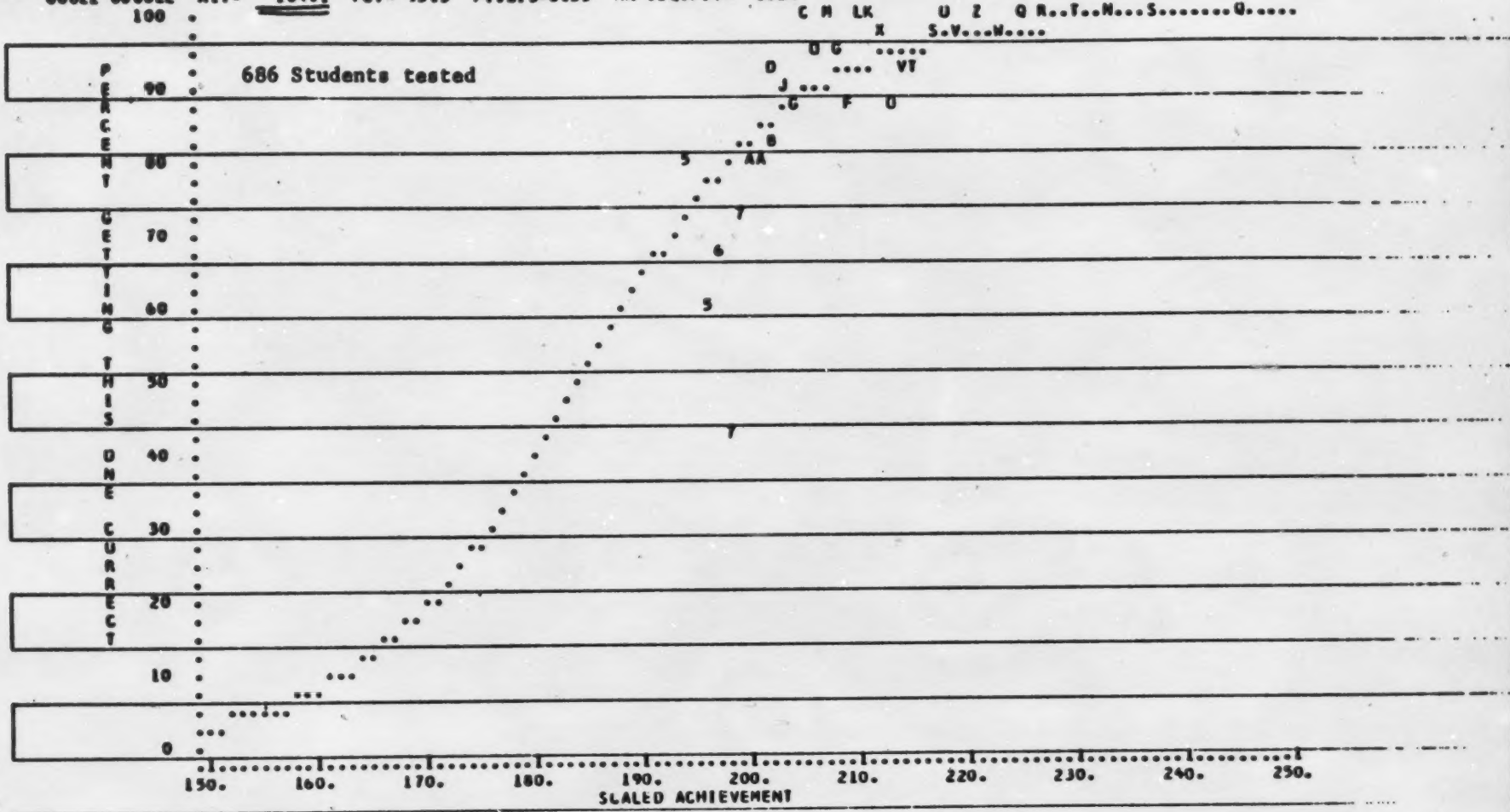


Figure 2C

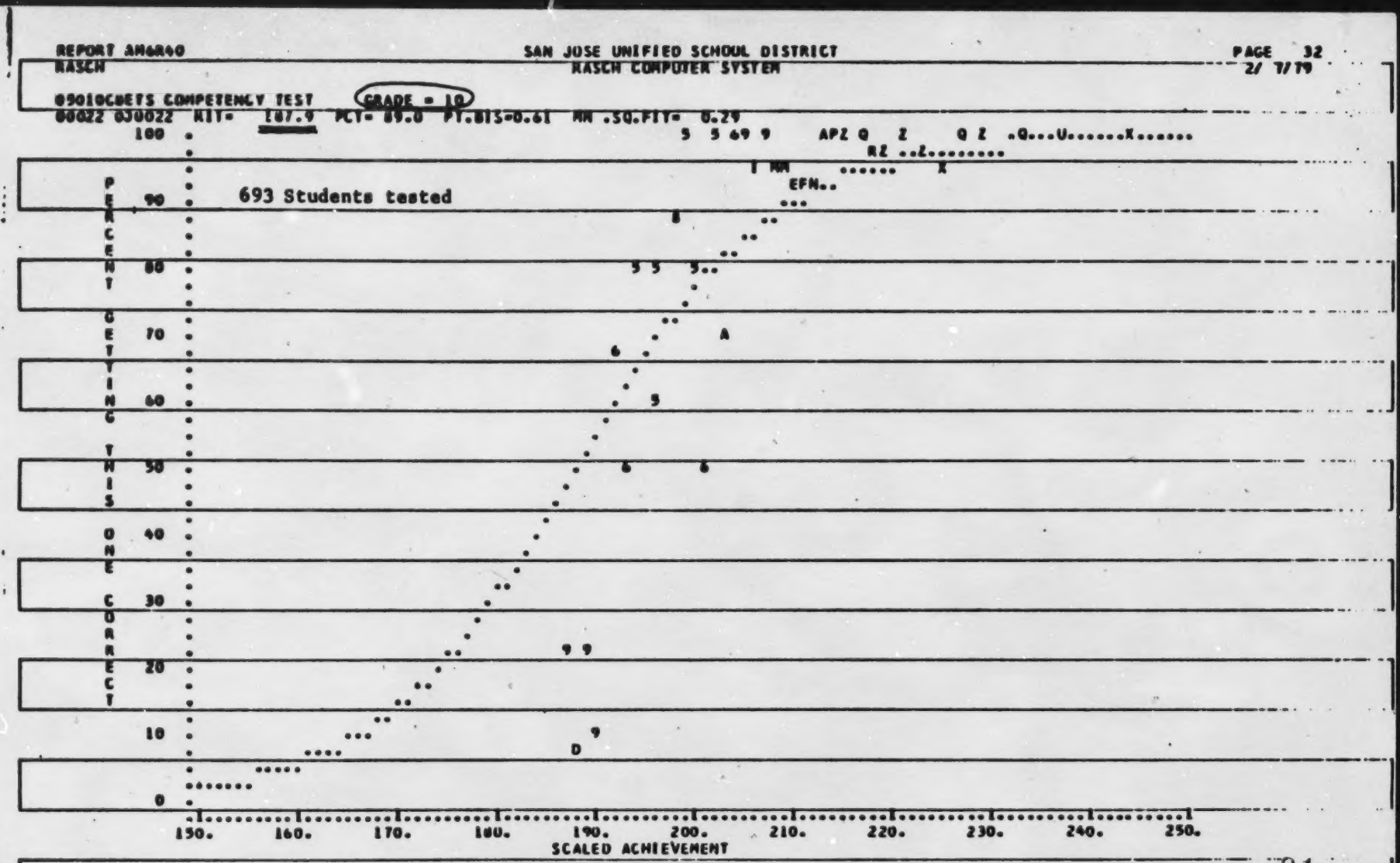
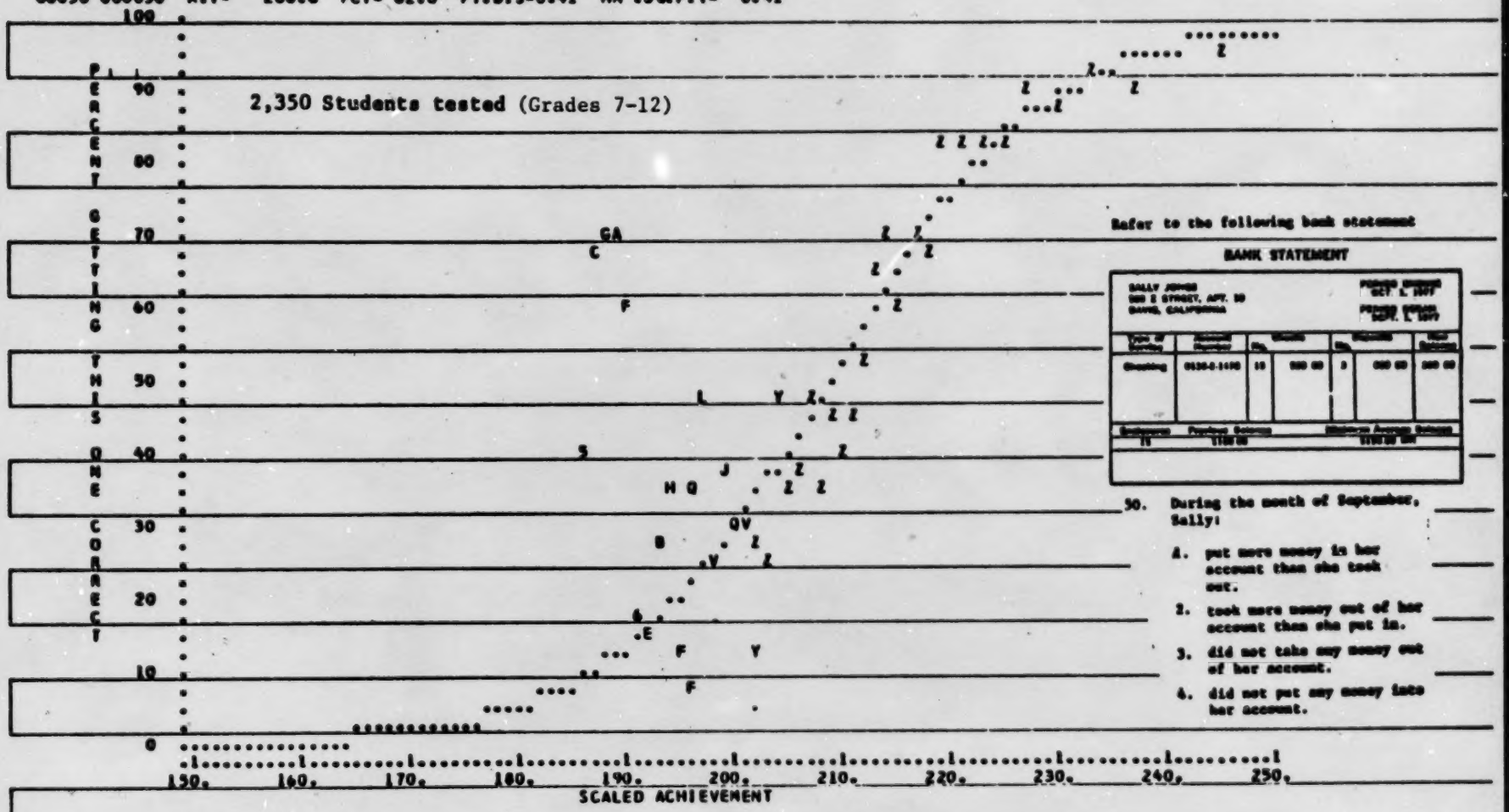

REPORT AM6R40
RASCH
SAN JOSE UNIFIED SCHOOL DISTRICT
RASCH COMPUTER SYSTEM
PAGE 32
2/ 7/79
0901OCBETS COMPETENCY TEST
GRADE = 10
00022 0J0022 MIT= 187.9 PCT= 89.0 PT.BIS=0.61 AM .SD.FIT= 0.29
693 Students tested
PERCENT GETTING THIS ONE CORRECT
100
90
80
70
60
50
40
30
20
10
0
150.
160.
170.
180.
190.
200.
210.
220.
230.
240.
250.
SCALED ACHIEVEMENT

Figure 2D

05010CBETS COMPETENCY TEST
00050 000050 RIT= 208.6 PCT= 62.0 PT.BIS=0.41 MN .SQ.FIT= 0.41



Refer to the following bank statement

BANK STATEMENT

50. During the month of September, Sally:

1. put more money in her account than she took out.
2. took more money out of her account than she put in.
3. did not take any money out of her account.
4. did not put any money into her account.

Figure 3A

09010CBETS COMPETENCY TEST GRADE = 08
00050 000050 RIT= 200.9 PCT= 57.3 PT.BIS=0.40 AN .50.FIT= 0.29

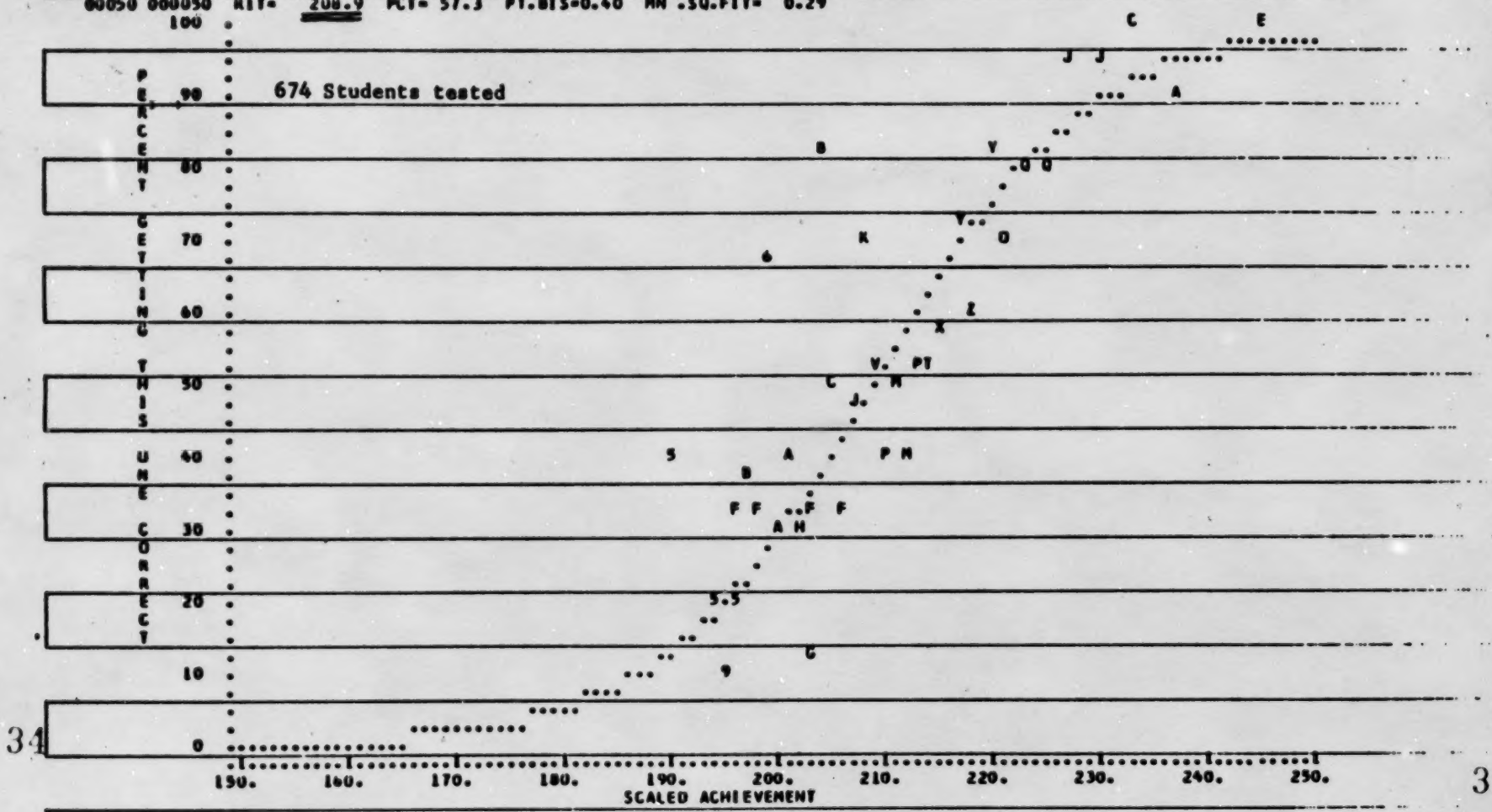


Figure 3B

09010CBETS CUMPETENCY TEST GRADE = 09
00050 000350 RIT= 209.3 PCT= 61.4 PT.BIS=0.49 AN .SU.FIT= 0.25

686 Students tested

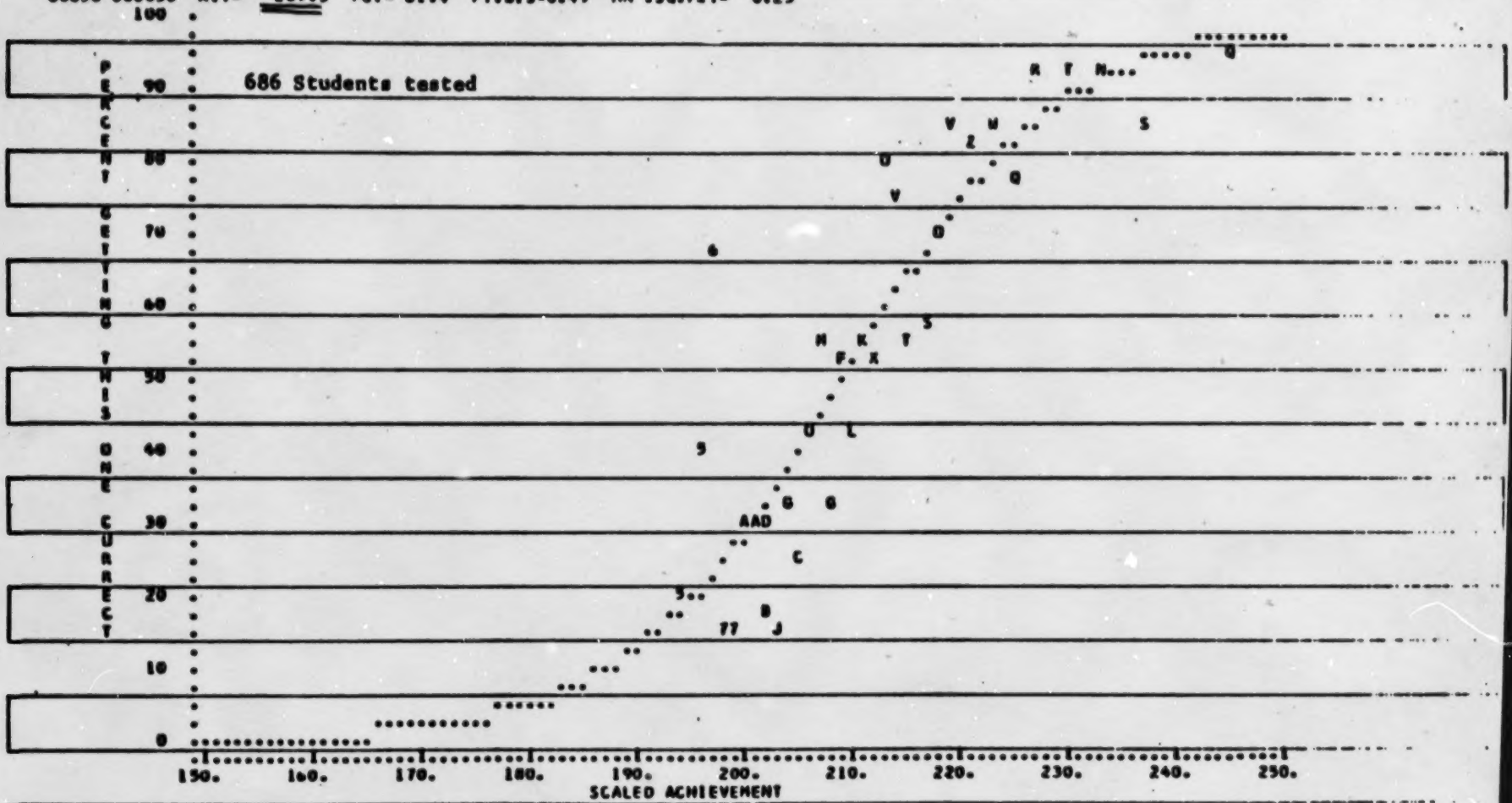


Figure 3C

05010CBETS COMPETENCY TEST GRADE = 10
00050 000050 AIT= 207.4 PCT= 64.5 PT.BIS=0.34 MN .SD.FIT= 1.49

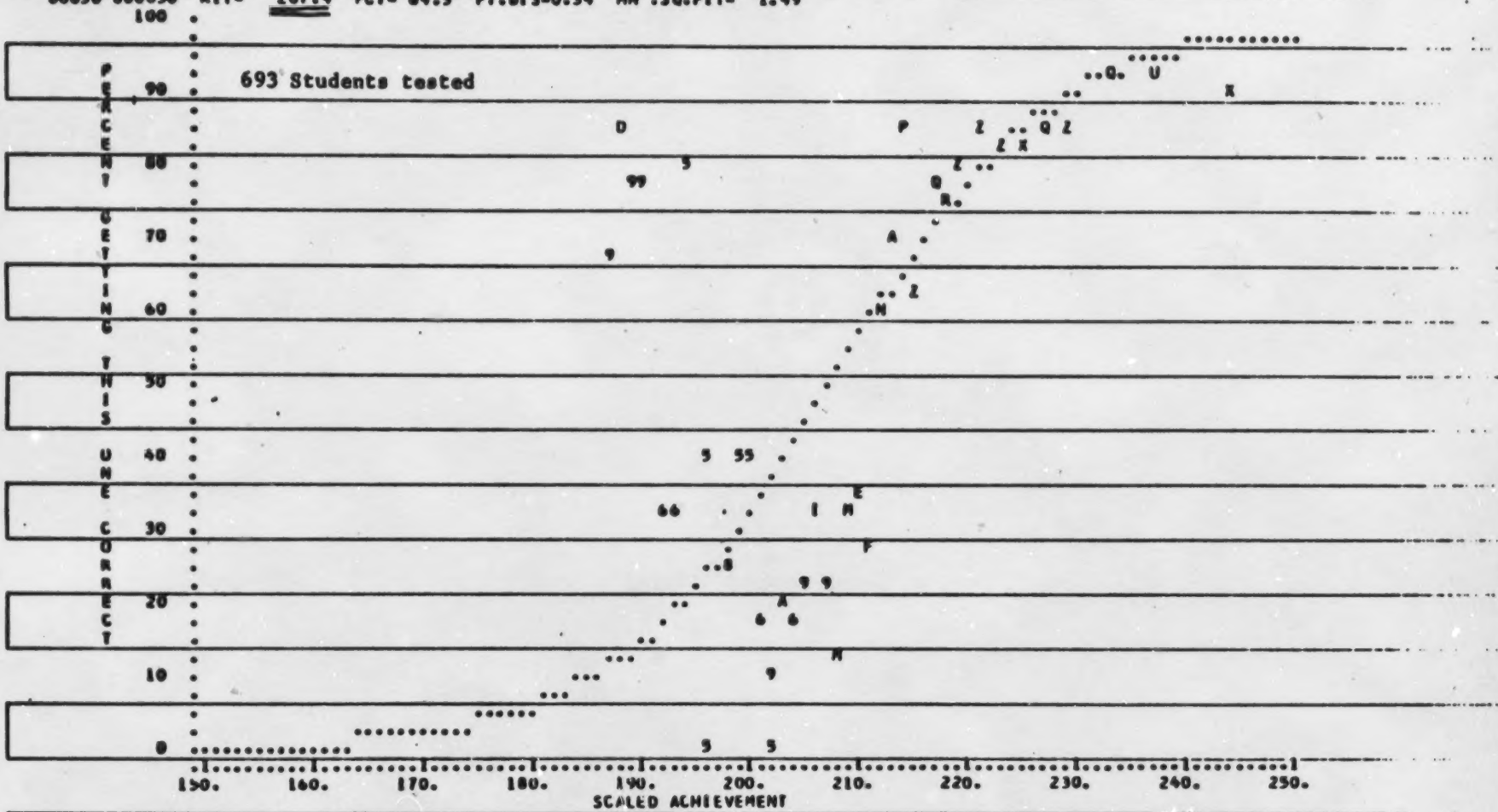


Figure 3D

APPENDIX A

SAN JOSE UNIFIED SCHOOL DISTRICT

COMPETENCY STATEMENTS
(AB 3408 and AB 65)
Board Adopted July 6, 1978

DEFINITIONS:

COMPETENCY - What a student can do (stated in general terms).
PERFORMANCE INDICATORS - How a student demonstrates performance in meeting the requirements of a competency.

MATHEMATICS COMPETENCIES

COMPETENCY 1

THE STUDENT CAN ADD, SUBTRACT, MULTIPLY, DIVIDE AND FIND THE PERCENT OF A NUMBER USING WHOLE NUMBERS, FRACTIONS AND DECIMALS.

PERFORMANCE INDICATORS

1.1 The student can add up to four whole numbers, each of which has four or fewer digits.

1.2 The student can find the difference between two whole numbers, each of which has four or fewer digits.

1.3 The student can multiply two whole numbers, each of which has three or fewer digits.

1.4 The student can divide a one or two digit whole number into another whole number showing the correct quotient and remainder.

1.5 The student can add up to four decimal numbers, each of which has four or fewer digits.

1.6 The student can find the difference between two decimal numbers, each of which has four or fewer digits.

1.7 The student can multiply two decimal numbers, each of which has three or fewer digits.

1.8 The student can divide a one or two digit decimal number into another decimal number rounding off the quotient to the nearest hundredth.

1.9 The student can add any two fractions (or mixed numbers less than ten) with denominators of two, three or four and express the sum as a mixed number in simplest terms.

1.10 The student can find the difference between any two fractions (or mixed numbers less than ten) with denominators of two, three or four and express the difference as a mixed number in simplest terms.

1.11 The student can multiply any two fractions (or mixed numbers less than ten) with denominators of two, three or four and express the product as a mixed number in simplest terms.

1.12 The student can divide one fraction (or mixed number less than ten) by another, both fractions with denominators of 2, 3 or 4, and express the quotient as a mixed number in simplest terms.

1.13 The student can find a one or two digit whole number percent of another whole number up to four digits in length and express the result rounded off to the nearest hundredth.

COMPETENCY 2

THE STUDENT CAN IDENTIFY AND USE THE TOOLS AND UNITS OF ENGLISH AND METRIC MEASUREMENT AND SOLVE PROBLEMS INVOLVING THEM.

PERFORMANCE INDICATORS

2.1 The student can choose the most reasonable unit of length, weight and capacity needed to measure representations of physical objects.

2.2 The student can use a thermometer to identify temperature to the nearest whole number.

2.3 The student can use a standard liquid measure to find the capacity of a container.

2.4 The student can identify the basic units used for measuring 1) length, 2) weight, 3) capacity and 4) temperature.

2.5 The student can correctly solve problems involving rate, time and distance.

2.6 The student can find the perimeter of a diagrammed figure.

2.7 The student can find the area of a diagrammed figure.

2.8 The student can identify the uses of different measurement instruments.

2.9 The student can use a ruler to express length.

2.10 The student can use a scale or balance to find the weight of a given object.

COMPETENCY 3

THE STUDENT CAN SOLVE CONSUMER PROBLEMS.

PERFORMANCE INDICATORS

3.1 The student can find the price of several identical items when the price of one item is given.

3.2 The student can find the price of one item when the price of several similar items is given.

3.3 The student can determine the total expenditure for all items on a shopping list when the price of each item is known.

3.4 The student can find the total cost of an item if given the amount of the down payment and the monthly payments needed to pay for the item.

3.5 The student can find the amount of tax and the total selling price of an article if given the price of the article and the rate of sales tax.

3.6 The student can determine the sales price of an item if given the original price and the rate of reduction.

3.7 The student can calculate the amount of change due after a given purchase.

COMPETENCY 4

THE STUDENT CAN DO WAGE AND TIME COMPUTATIONS.

PERFORMANCE INDICATORS

4.1 The student can compute the elapsed time between the beginning and ending times of an activity.

4.2 The student can find the amount of pay a worker should receive when the rate of pay per hour and the number of hours worked are given.

4.3 The student can determine the amount of take-home pay when the gross salary and itemized deductions are given.

COMPETENCY 5

THE STUDENT CAN USE BASIC MONEY MANAGEMENT SKILLS.

PERFORMANCE INDICATORS

5.1 The student can determine the amount of money available for other expenses when the fixed monthly expenses and the amount of take-home pay are given.

5.2 The student can determine the total annual cost and the average monthly cost of an automobile when the individual operating costs are given.

5.3 The student can determine the difference in total cost between a cash purchase and a credit purchase involving installment payments, when the original cost of the item and the interest charges are known.

5.4 The student can find the interest on a specified amount of savings for a specified time if the annual interest rate is known.

5.5 The student can find the new balance for a checking account when the beginning balance and the transactions that have subsequently taken place are known.

5.6 The student can recognize and demonstrate an understanding of checking account statements.

5.7 The student can understand and extract pertinent information from an earnings statement.

COMPETENCY 6

THE STUDENT CAN INTERPRET NUMERICAL INFORMATION PRESENTED IN MAPS, SIMPLE GRAPHS, TABLES, AND LISTS OF RELATED DATA.

PERFORMANCE INDICATORS

6.1 The student can interpret data presented in a graph.

6.2 The student can interpret and/or use numerical information presented on a map.

6.3 The student can understand and extract pertinent information from a table.

6.4 The student can compute the average of a list of related data.

SAN JOSE UNIFIED SCHOOL DISTRICT

COMPETENCY STATEMENTS
(AB 3408 and AB 65)
Board Adopted July 6, 1978

DEFINITIONS:

COMPETENCY - What a student can do (stated in general terms).
PERFORMANCE INDICATORS - How a student demonstrates performance in meeting the requirements of a competency.

READING COMPETENCIES

COMPETENCY 1

STUDENT CAN READ AND UNDERSTAND WRITTEN DIRECTION.

PERFORMANCE INDICATORS

1.1 The student can read and interpret information presented on a label or labels.

1.2 The student can read and demonstrate understanding of written directions presented in instructional manuals.

1.3 The student can read and demonstrate understanding of the written directions of a recipe.

1.4 The student can read and use the information on a sample clothing tag.

COMPETENCY 2

STUDENT CAN READ AND UNDERSTAND PRINTED SIGNS, MAPS, CHARTS, AND SCHEDULES.

PERFORMANCE INDICATORS

2.1 The student can indicate the proper meaning of a printed sign.

2.2 The student can interpret information presented on a map.

2.3 The student can interpret information presented on a chart or table.

2.4 The student can interpret information on a printed bus, plane, or train schedule.

COMPETENCY 3

STUDENT CAN READ TO DETERMINE DETAILS, THE ORDER OF EVENTS, MAIN IDEAS, AND WORD MEANINGS IN CONTEXT.

PERFORMANCE INDICATORS

3.1 The student can read a selection for details and answer questions about its content.

3.2 The student can identify the main idea of a reading selection.

3.3 The student can indicate the order of events in a reading selection.

3.4 The student can identify the meaning of a word in the context of a sentence.

COMPETENCY 4

STUDENT CAN READ AND UNDERSTAND A VARIETY OF FORMS.

PERFORMANCE INDICATORS

4.1 The student can extract pertinent information from an application for a Social Security number.

4.2 The student can read and extract pertinent information from sample voter registration forms and information about registration.

4.3 The student can read and extract pertinent information from driver-related information including traffic regulations, vehicle accident reports and vehicle codes to answer questions similar to those found in tests for the driver's license.

COMPETENCY 5

STUDENT CAN USE LIBRARY RESOURCES AND OTHER REFERENCE MATERIALS.

PERFORMANCE INDICATORS

5.1 The student can define words and understand other kinds of information presented in a dictionary.

5.2 The student can indicate the author, title, subject, and call number of the book from information found on a sample library catalog card.

5.3 The student can extract pertinent information from a telephone directory.

5.4 The student can use the Table of Abbreviations of a textbook or reference book.

5.5 The student can use a sample index to extract appropriate information.

5.6 Student can extract pertinent information from a sample table of contents.

5.7 Student can organize data in an alphabetically-ordered file.

COMPETENCY 6

STUDENT CAN EXTRACT PERTINENT INFORMATION FROM A VARIETY OF WRITTEN MATERIALS.

PERFORMANCE INDICATORS

6.1 The student can distinguish fact from opinion in related statements.

6.2 The student can analyze statements made in advertisements.

6.3 The student can read newspapers and magazines for information related to daily life.

6.4 The student can demonstrate an understanding of the contents of a personal letter.

6.5 The student can demonstrate an understanding of the contents of a business letter.

6.6 The student can demonstrate an understanding of cursive writing.

SAN JOSE UNIFIED SCHOOL DISTRICT

COMPETENCY STATEMENTS
(AB 3408 and AB 65)
Board Adopted July 6, 1978

DEFINITIONS:

COMPETENCY - What a student can do (stated in general terms).
PERFORMANCE INDICATORS - How a student demonstrates performance in meeting the requirements of a competency.

WRITING COMPETENCIES

COMPETENCY 1

STUDENT CAN WRITE A PARAGRAPH OR PARAGRAPHS THAT DEMONSTRATE PROFICIENCY IN THE FOLLOWING AREAS:

1. Complete sentences
2. Correct spelling
3. Correct punctuation
4. Legibility
5. Purpose
6. Sequence
7. Clarity

PERFORMANCE INDICATORS

1.1 The student can describe an incident or scene.

1.2 The student can write directions for something s/he knows how to make or do.

1.3 The student can write a business letter applying for a job or inquiring about a product or service.

COMPETENCY 2

STUDENT CAN FILL OUT FORMS.

PERFORMANCE INDICATORS

2.1 The student can fill out a check.

2.2 The student can fill out a blank job application.

2.3 The student can fill out a blank Social Security form.

2.4 The student can fill out a simplified tax form.

2.5 The student can fill out a credit application.

REFERENCES

Begle, E.G. Introductory Calculus with Analytic Geometry, Henry Hold and C o. 1954, 42-48.

Birnbaum B. in Hambleton R.K. et al Developments in Latent Trait Theory, Review of Educational Research, Fall 1978, 48, 490-491.

Choppin, B., Item Banking and the Monitoring of Achievement: an Introductory Paper, National Foundation for Educational Research in England and Wales, 1978, 1-40.

Forster, F. and Ingebo, G. Application of the Rasch Model in School Settings, paper presented at annual meeting of California Educational Research Association, San Diego, November 1979.

Forster, F. and Ingebo, G. Rasch Model Monograph Series, December 1978, 1, 1-34 Northwest Evaluation Association.

Hambleton, R.K. et al, Developments in Latent Trait Theory: Models, Technical Issues, and Applications in Review of Educational Research, Fall 1978, 48, 467-503.

McClung, M.S., Competency Testing: Potential for Discrimination, Clearing House Review, September 1977, 439-442.

Moriel, J. and Davis, J.N. How to Develop Equivalent Competency Tests: Using the Rasch Measurement Model, paper presented at annual meeting of California Educational Research Association, San Diego, November 1979.

Mullis, I.V.S. and Oldefendt, S.J. and Phillips, D.L. What Students Know and Can Do, Profiles of three age groups. National Assessment of Educational Progress, Denver, Colorado, March 1977.

Ryan, J.P. and Hamm, D.W., An Introduction to the Rasch Latent Trait Psychometric Model, paper presented at AERA, Toronto, April 1978.

San Jose Unified School District, Competency Statements for High School Graduation, Board of Education proceedings, 7/6/78.

Wright, B.D., Solving Measurement Problems with the Rasch Model, in Journal of Educational Measurement, Summer 1977, 14, 97-114.

Zieby, M.J. and Fremer, J., Guidelines for Developing and Using Minimum Competency Tests, paper presented at AERA annual meeting, San Francisco, April 1979.

Zieby, M.J. and Livingston, S.A., Manual for Setting Standards on the Basic Skills Assessment Tests, Educational Testing Service, Princeton, N.J. 1977.